Memorandum

To: Internal Correspondence - Conflict Probe Assessment Team (CPAT)

CC: Ron Dloughy, *Aerospace Engineering*

From: Mike Paglione, *FAA ACB-330*

Date: 5/28/2002

Re: **ACB-330 Informal Notes on Comparing Two Runs or Two Decision Support Tools Conflict Alert Predictions Given the Same Traffic Scenario, Version 1**

Scope

ACB-330 is evaluating the accuracy sensitivity of URET's predictions to specified weather forecast error. Also pending a decision by AOZ, ACB-330 will be comparing the predictions of both URET and CTAS Direct-To (D2). In both cases, there is a need for a methodology of comparing the conflict predictions of two Decision Support Tools (DST) or the same DST under varying conditions.

This memo will document ACB-330's initial notes on comparing two DST conflict prediction results. The memo will provide the foundation for software tool development to perform the comparison operations as well as manual spreadsheet calculations.

Objective of Analysis

The objective of the analysis is to determine the difference and/or similarity of the conflict predictions between DST 1/Run 1 to DST 2/Run 2. For the weather forecast sensitivity study, the goal is slightly more specific, since the focus is on the difference or similarity between the control run to the treatment run. The focus for the weather study will be on the impact of the induced weather forecast error on the treatment runs.

The conflict predictions can include:

- Predictions for aircraft to aircraft encounters or conflicts[1] with and without flight plan adherence
- Predictions for aircraft to special use airspace encounters and conflicts with and without flight plan adherence

URET notifications may include red, yellow, both red and yellow, blue, or muted alerts.

---

[1] Conflicts between aircraft and special use airspace are in regard to standard legal separation (e.g. 5 nautical miles and 1000/2000 feet vertically). Encounters are assumed to be at larger separations both horizontally and vertically (e.g. 30 nautical miles and 4000/5000 feet vertically).

Initial Analysis Assumptions

1. It is assumed that both decision support tools (DSTs) or runs will be input with the same set of air traffic scenarios for the same adaptation chart cycle. Therefore, the ground truth information will be the same in evaluation of both DST's predictions. The Host Computer System (HCS) messages are mainly used to determine the ground truth (i.e. actual position of the aircraft, intent of the aircraft, and any conflicts between aircraft) extracted from the scenario.

2. Since the focus of the study is to examine the differences of two DST's or run's predictions not measuring their absolute accuracy, one well-defined scenario input through both DSTs is assumed sufficient. These scenarios will be constructed from actual traffic data and only modified by time to induce a reasonable number of conflict/encounters. This approach as been used very successfully in both the formal accuracy testing of URET Core Capability Limited Deployment (URET CCLD) and the Risk Reduction Program for URET CCLD[2].

3. Not only the same traffic input data is provided into the DST(s) as described in assumptions 1 and 2 above, but the evaluation of the individual conflict predictions of these runs will be determined under the same rules. For example, the calculation of valid, missed, and false alerts for a given comparison will include adherence or not include adherence rules. For the given comparison, both runs must apply the same adherence rules. Without consistency in the applied rules, the analysis would be confounded. However, the analysis can be repeated and evaluated independently under varying rules.

Operational Metrics for Conflict Predictions

The conflict prediction accuracy metrics describe two fundamental events: a conflict occurs and/or an alert is predicted. These events, which are not mutually exclusive, have four possible outcomes (see Table 1). The conflict accuracy metrics measure the two fundamental error outcomes: missed alert and false alert. This is explained in detail in references [6], [7], and [8].

	CONFLICT OCCURS	CONFLICT DOES NOT OCCUR
ALERT	DST predicts conflict and it occurs (V -- valid alerts)	DST predicts conflict and it does not occur (F -- false alert)
NO ALERT	DST does not predict conflict and it occurs (M -- missed alert)	DST does not predict conflict and it does not occur (NC -- correct no-calls)
Total Number of Alerts	Total Number of Conflicts	Total Number of Non-Conflicts (Encounters that did not have conflicts)

Table 1: DST Alert and Conflict Event Combinations

[2] URET CCLD Risk Reduction Program involves informal accuracy tests to examine issues prior to deployment to the various sites.

With two collocated DSTs or two runs from the same DST presenting their own conflict predictions (alerts), the first column in Table 1 expands to consider all combinations of intersection and union of these events. This is illustrated in Table 2. Once again, these combinations assume both DST's are using the same definition of an alert and the comparisons are based on the same ground truth conflict/non-conflict event. Actually, both are critical elements for this type of study.

	CONFLICT OCCURS	CONFLICT DOES NOT OCCUR
ALERT by both Runs A &B or DST A & B	**Both predicts conflict and it occurs (V_{A1}=V_{B1}-- valid alerts both)**	**Both predicts conflict and it does not occur (F_{A1}=F_{B1}-- false alert both)**
ALERT by A and	A predicts conflict and it occurs (V_{A2} -- valid alerts by A only)	A predicts conflict and it does not occur (F_{A2} -- false alert by A only)
not B	B does not predict conflict and it occurs (M_{B2} -- missed alert by B only)	B does not predict conflict and it does not occur (NC_B -- correct no-calls by B only)
ALERT by B and	B predicts conflict and it occurs (V_{B2} -- valid alerts by B only)	B predicts conflict and it does not occur (F_{B2} -- false alert by B only)
not A	A does not predict conflict and it occurs (M_{A2} -- missed alert by A only)	A does not predict conflict and it does not occur (NC_A -- correct no-calls by A only)
NO ALERT by both Runs A &B or DSTs	**Both do not predict conflict and it occurs (M_{A1}= M_{B1} -- missed alert by both**	**Both do not predict conflict and it does not occur (NC -- correct no-calls by both)**
Total Number of Alerts for each/both[3]	Total Number of Conflicts *(Same for both DSTs or Runs!)*	Total Number of Non-Conflicts (Encounters that did not have conflicts; *Same for both DSTs!*)

Table 2: Two DSTs or Two Runs of the Same DST Alert and Conflict Event Combinations

The above analysis assumed standard legal separation conflicts when referring to conflicts in Table 1 and 2, but the analysis could also include repeat runs considering expanded conflicts (a.k.a. encounters) of various separations to examine events of interest to the controller that will not necessarily cause a legal separation violation. Additional advanced metrics could be examined as well, such as the "sharpness" of the prediction curve, see references [6], [7], and [8] for details.

The situations described in Table 2 are thorough but not exhaustive. Table 2 does not include all the possible events when you consider the rules applied to determine the missed, false, and valid alerts. Specifically, for missed and valid alert combinations flight plan adherence could exclude certain missed alerts and thus conflicts under Run A (or DST A) and not under Run B (or DST B). In other words, Run B (or DST B) could have successfully predicted the particular conflict resulting in a valid alert, while Run A (or DST A) did not. For Run A under this situation, the adherence rule allowed the missed alert to be discarded. Other discard rules are applied as well, particularly for false alerts. Therefore, Table 3 expands the situations as described in Table 2 to include all discard cases.

[3] For the total number of alerts one could take the individual total counts of each Run's / DST's alerts and their union and intersection.

	CONFLICT OCCURS	CONFLICT DOES NOT OCCUR
ALERT by both Runs A &B or DST A & B	**Both predicts conflict and it occurs** **(V_{A1}=V_{B1}-- valid alerts both)**	**Both predicts conflict and it does not occur** **(F_{A1}=F_{B1}-- false alert both)**
ALERT by A and	A predicts conflict and it occurs (V_{A2} -- valid alerts by A only)	A predicts conflict and it does not occur (F_{A2} -- false alert by A only)
not B	B does not predict conflict and it occurs (M_{B2} -- missed alert by B only)	B does not predict conflict and it does not occur (NC_B -- correct no-calls/discards by B only)
ALERT by A and	A predicts conflict and it occurs (V_{A3} -- valid alerts by A only)	A predicts conflict and it does not occur (** F_{A2} Continued **)
B ALERT or non-ALERT is discarded	B does not predict conflict correctly but is discarded ($Discard_B$ -- B discards only)	B does not predict conflict correctly but is discarded (** NC_B Continued **)
ALERT by B and	B predicts conflict and it occurs (V_{B2} -- valid alerts by B only)	B predicts conflict and it does not occur (F_{B2} -- false alert by B only)
not A	A does not predict conflict and it occurs (M_{A2} -- missed alert by B only)	A does not predict conflict and it does not occur (NC_A -- correct no-calls/discards by A only)
ALERT by B and	B predicts conflict and it occurs (V_{B3} -- valid alerts by B only)	B predicts conflict and it does not occur (** F_{B2} Continued **)
A ALERT or non-ALERT is discarded	A does not predict conflict correctly but is discarded ($Discard_A$ -- A discards only)	A does not predict conflict correctly but is discarded (** NC_A Continued **)
NO ALERT by both Runs A &B or DSTs	**Both do not predict conflict and it occurs** **(M_{A1}= M_{B1} -- missed alert by both**	**Both do not predict conflict and it does not occur** **(NC -- correct no-calls by both)**
Total Number of Alerts for each/both[4]	Total Number of Conflicts *(Same for both DSTs or Runs!)*	Total Number of Non-Conflicts (Encounters that did not have conflicts; *Same for both DSTs!*)

Table 3: Two DST Alert and Conflict Event Combinations With Discard Events

[4] For the total number of alerts one could take the union of each Run's / DST's alerts by adding both sets of valid alerts and false alerts subtracted by the common valid and false alerts.

The verifiable conflicts for Run A or Run B are slightly different due to the potential for discarding. For discarding missed alerts, the only rule that could possibly apply is the flight plan adherence of the true conflict. The conflict would have to have an adherence age of beyond a parameter time at the start of conflict (e.g. 13 minutes). Valid alerts of conflicts with adherence age less than the threshold time are correct, while missed alerts with the same adherence age can be discarded. This prevents penalizing a DST from predicting alerts correctly even if the input intent of the flights is in error. Thus, the number of verifiable conflicts for a given run or DST is the composite of the valid and missed alerts. Equation 1 is the total verifiable conflicts for Run A (or DST A), while Equation 2 is the same for Run B (or DST B). Equation 3 lists the quantity of all the verifiable conflicts for both.

$$C_A = V_A + M_A = V_{A1} + V_{A2} + V_{A3} + M_{A1} + M_{A2} \tag{1}$$

$$C_B = V_B + M_B = V_{B1} + V_{B2} + V_{B3} + M_{B1} + M_{B2} \tag{2}$$

$$C_{ALL} = C_A \cup C_B = C_A + C_B - (C_A \cap C_B)$$

$$C_{ALL} = C_A + C_B - V_{B1} - V_{B2} - M_B \tag{3}$$

Table 4 summarizes the event count variables in Table 3. Each column, except the first referring to the total conflict counts per run, represents variables that are equivalent. For example, V_{A1} is equal to V_{B1}.

Conflicts	Common Valid Alerts	Valid A and Missed B	Valid A and Discard B[5]	Common Missed Alerts	Missed A and Valid B	Discard A And Valid B[6]
C_A	V_{A1}	V_{A2}	V_{A3}	M_{A1}	M_{A2}	$Discard_A$
C_B	V_{B1}	M_{B2}	$Discard_B$	M_{B1}	V_{B2}	V_{B3}

Table 4: Summary of Event Count Variables

To compare the two runs (or DSTs), the difference in missed alert probability is of interest. Equation 4 is the missed alert probability for Run A and Equation 5 is the same for Run B.

$$\text{Run A Probability of Missed Alert} = \frac{M_A}{C_A} \tag{4}$$

$$\text{Run B Probability of Missed Alert} = \frac{M_B}{C_B} \tag{5}$$

$$\text{Missed Alert Probability Difference} = \frac{M_A C_B - M_B C_A}{C_A C_B}$$

$$\text{Missed Alert Probability Difference} = \frac{M_A V_B - M_B V_A}{C_A C_B} \tag{6}$$

[5] V_{A3} are valid alerts only if flight plan adherence is used to discard conflicts in B.
[6] V_{B3} are valid alerts only if flight plan adherence is used to discard conflicts in A.

Equation 6 is the general equation of the difference between runs (or DSTs) in missed alert probability. If the adherence rule is not applied, there will be no discarded conflicts in both runs and Equation 6 is simplified to the following Equation 7.

$$\text{If no adherence rule, then } C_A = C_B = C \text{ and } V_{A3} = V_{B3} = 0.$$

$$\text{Missed Alert Probability Difference} = \frac{M_{A2} - V_{A2}}{C} = \frac{V_{B2} - M_{B2}}{C} \quad (7)$$

Analogous to the missed alert probabilities, false alert probabilities can be examined also. The difference is the adherence rule does discard some false alerts, but there are several other rules which can allow the discard of false alerts. These are listed in Table 5. The conditional false alert probabilities are listed in Equation 8 and 9 for Runs A and B, respectively. The difference in false alert probabilities between runs is listed in Equation 10.

Reason Code	Description
NO_TRK_FA_DISCARD	No post processed track a predicted conflict start time so discard
NO_ADHER_FA_DISCARD	Out of adherence at predicted conflict start time so discard
CLR_FA_DISCARD	Retracted FA assigned by an ATC clearance so discard
CFL_FA_DISCARD	FA notified beyond last conflict actual start time so discard

Table 5: Discard Events for False Alerts[7]

$$\text{Run A Probability of False Alert} = \frac{F_A}{A_A} \quad (8)$$

$$\text{Run B Probability of Missed Alert} = \frac{F_B}{A_B} \quad (9)$$

$$\text{False Alert Probability Difference} = \frac{F_A A_B - F_B A_A}{A_A A_B}$$

$$\text{False Alert Probability Difference} = \frac{F_A V_B - F_B V_A}{A_A A_B} \quad (10)$$

With A_A being the quantity of alerts for Run A and A_B being the quantity of alerts for Run B, it is also necessary to find the total quantity of alerts for both runs, which is analogous to the total number of conflicts expressed in Equation 3. Equation 11 expresses the number of Run A alerts and Equation 12 lists the number of Run B alerts. Equation 13 expresses the union of these two runs or the total number of alerts for both runs.

[7] These discard rules were originally developed for the URET CCLD Formal Accuracy Test Program but will be applied to this analysis as well.

$$A_A = V_A + F_A = V_{A1} + V_{A2} + V_{A3} + F_{A1} + F_{A2} \quad (11)$$

$$A_B = V_B + F_B = V_{B1} + V_{B2} + V_{B3} + F_{B1} + F_{B2} \quad (12)$$

$$A_{ALL} = A_A \cup A_B = A_A + A_B - (A_A \cap A_B)$$

$$A_{ALL} = A_A + A_B - V_{B1} - F_{B1} = A_A + A_B - V_{A1} - F_{A1} \quad (13)$$

Besides the missed and false alert differences, there are other quantities of interest. These additional comparison probabilities are summarized in Table 6.

Equation	Description	Equation Number
$\frac{M_{A1} \Leftrightarrow M_{B1}}{C_{All}}$	Common missed alert probability, that is the probability that both runs had missed the conflict	14
$\frac{V_{A1} \Leftrightarrow V_{B1}}{C_{All}}$	Common valid alert probability, that is the probability that both runs had correctly predicted the conflict	15
$\frac{V_{A2} \Leftrightarrow M_{B2}}{C_{All}}$	Probability that Run A correctly called the conflict while Run B missed the conflict	16
$\frac{V_{B2} \Leftrightarrow M_{A2}}{C_{All}}$	Probability that Run B correctly called the conflict while Run A missed the conflict	17
$\frac{F_{A1} \Leftrightarrow F_{B1}}{A_{All}}$	Common conditional false alert probability, that is the probability that both runs had a falsely predicted a conflict	18

It is also necessary to examine various parameter distributions of valid alerts. These include:

- Conflict notification start time and end time
- Predicted conflict start time and end time
- Warning time, defined as notification start time subtracted by actual conflict start time (i.e. notif_start_time – ACST)
- Sector the alert is presented versus the sector where the conflict started

Although to varying degrees, ACB-330 suspects all these parameters would be of interest to DST users if not equivalent or very close. Distribution plots, point statistics, and statistical tests could be employed to compare these metrics as previously done for trajectory accuracy in references [1], [2], and [3].

In a similar manner, statistical tests could be employed to test the hypothesis that the missed and false alert probabilities from Run A (or DST A) are equivalent to Run B (or DST B). This will be examined in Version 2 of this memorandum. Version 3 will present the sharpness and sharpness bias metrics as defined in references [7] and [8]. Appendix A presents evaluation codes for a software tool that compares Run A conflict alert predictions against Run B.

References

1. *Trajectory Prediction Accuracy Report: URET/CTAS* (DOT/FAA/CT-TN99/10), Paglione, M. et. al., WJHTC/ACT-250, May 1999.

2. "A Generic Sampling Technique for Measuring Aircraft Trajectory Prediction Accuracy," Cale, M., Liu, S., Oaks, B., Ryan, H., Summerill, S., 4th U.S.A./Europe Air Traffic Management R&D Seminar, December 3, 2001 (pending).

3. Briefing at Performance Technical Interchange Meeting at Lockheed Martin Air Traffic Management, "Quick Example of Benefits of Pairing Trajectory Accuracy Data," Paglione, M., July 24, 2000.

4. Devore, Jay L., Probability and Statistics for Engineering and the Sciences, Second Edition, 1987.

5. Montgomery, Douglas C., Design and Analysis of Experiments, Fourth Edition, 1997.

6. *URET Conflict Prediction Accuracy Report* (DOT/FAA/CT-TN98/8), Cale, M., Paglione, M., Ryan, Dr. H., Timoteo, D., Oaks, R., WJHTC/ACT-250, April 1998.

7. "Generic Metrics for the Conflict Probe Tools Developed for Free Flight", Paglione, M., Ryan, Dr. H., Kazunas, S., Cale, M., 42nd Annual Air Traffic Control Association Conference Proceedings, Washington, DC, Fall 1997.

8. "Generic Metrics for the Estimation of the Prediction Accuracy of Aircraft to Aircraft Conflicts by a Strategic Conflict Probe Tool", Paglione, M., Cale, M., Ryan, Dr. H., Air Traffic Quarterly, Air Traffic Control Association, Volume 7(3) 147-165, 1999.

APPENDIX A: Event Evaluation Codes

Event	Evaluation Code	Description
V_{A1} or V_{B1}	SAME_VA	Both runs have valid alerts for the same conflict
M_{A1} or M_{A1}	SAME_MA	Both runs have missed alerts for the same conflict
F_{A1} or F_{B1}	SAME_FA	Both runs have false alerts for the same encounter
V_{A2} or M_{B2}	VA_MA	Run A has a valid alert and Run B has a missed alert for the same conflict
M_{A2} or V_{B2}	MA_VA	Run A has a missed alert and Run B has a valid alert for the same conflict
V_{A3} or $Discard_B$	VA_DISCARD	Run A has a valid alert while Run B discards the conflict
$Discard_A$ or V_{B3}	DISCARD_VA	Run A discards the conflict while Run B has a valid alert
F_{A2} or NC_B	FA_NC	Run A has a false alert while Run B either has no prediction or discards the alert for the same encounter
NC_A or F_{B2}	NC_FA	Run A either has no prediction or discards the alert while Run B has a false alert for the same encounter

Table A.1: Conflict Prediction Comparison Program Evaluation Codes